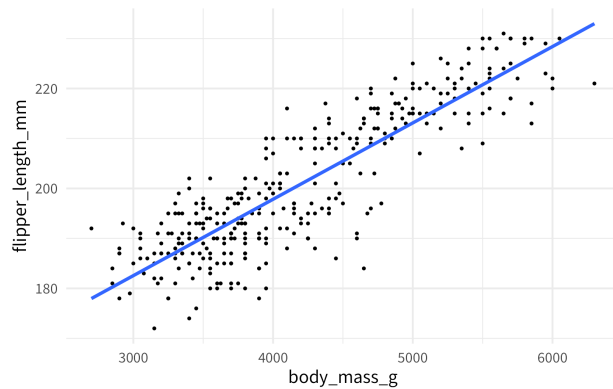


NOTES 07: LINEAR REGRESSION

Stat 120 | Fall 2025

Prof Amanda Luby



$$\widehat{\text{flipper_length_mm}} = 136.73 + 0.02(\text{body_mass_g}) \quad (1)$$

- Intercept (b_0): where does the line cross the y-axis? What is the *prediction* for $x = 0$?
- Slope (b_1): For a 1-unit increase in x , what is the change in the *prediction* for y ?

Observation

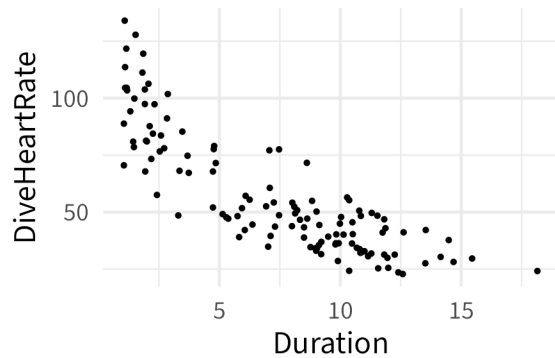
Prediction

Residual

Example: I grab a penguin with a flipper length of 193mm and a body mass of 3475g. What is the **observation**, **prediction**, and **residual** in this case?

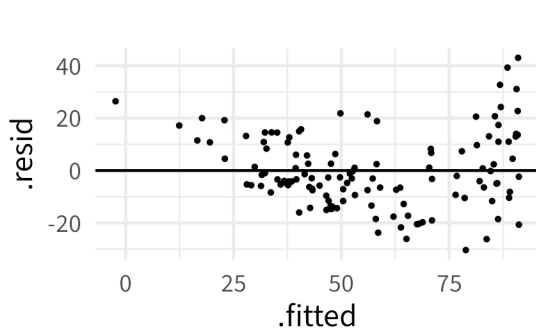
Example: Emperor penguins routinely make dives of 5-12 minutes, with the longest recorded dive over 27 minutes (!!). The rate of oxygen depletion is primarily determined by the penguin's heart rate. DiveHeartRate is the bird's heart rate in beats per minute, Depth is the depth of the dive (in meters),

Duration is the duration of the dive in minutes, and Bird is an ID variable indicating which penguin made the dive.

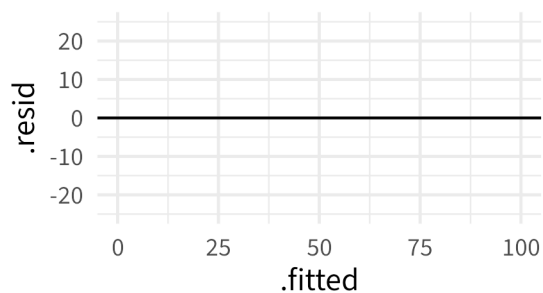


It can be helpful to look at a *residual scatterplot* when checking for linearity and outliers.

```
dives_lm = lm(DiveHeartRate ~ Duration, data = penguin_dives)
dives_aug = augment(dives_lm)
p1 = ggplot(dives_aug, aes(x = .fitted, y = .resid)) +
  geom_point(size = .5) +
  geom_hline(yintercept = 0)
```



What would an ideal residual plot look like?



Below is the slope and intercept for the transformed model. Write out the linear regression equation and provide an in-context interpretation of the intercept and slope.

(Intercept)	Duration
-0.007926	-0.001817